The Validity of Instruments Revisited

Daniel Berkowitz *
University of Pittsburgh

Mehmet Caner**
North Carolina State University

Ying Fang***
Xiamen University

September 22, 2011

Abstract

This paper shows how valid inferences can be made when an instrumental variable does not perfectly satisfy the orthogonality condition. When there is a mild violation of the orthogonality condition, the Anderson-Rubin (1949) is oversized. In order to correct this problem, the fractionally resampled Anderson-Rubin test is derived by modifying Wu's (1990) resampling technique. We select half of the sample when resampling and obtain valid but conservative critical values. Simulations show that our technique performs well even with moderate to large violation of exogeneity when there is a finite sample correction for the block size choice.

Keywords: Berry-Esseen Bound, Finite Sample of Random Variables, Near-Exogeneity

JEL Code: C3, C13, C30.

*Daniel Berkowitz: Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, email:dmberk@pitt.edu. ** Mehmet Caner, Corresponding Author: Department of Economics, 4168 Nelson Hall, Raleigh, NC 27695. email: mcaner@ncsu.edu, tel: 1-919-513-0853, fax: 1-919-513-0854. *** Ying Fang: WISE, Xiamen University, China email:yifst1@gmail.com.

1 Introduction

Instrumental variable estimation is one of the most widely used methods in economics. Valid instruments must be relevant and exogenous. Regarding the relevance of instruments, there has recently been a growing interest in the asymptotics of weak instruments. One of the most widely used test statistics in that research line is the Anderson-Rubin (1949) test (for herein denoted the AR test). The AR test statistic can be used when instruments are weak as shown by Stock and Wright (2000). Regarding exogeneity, however, there is a growing consensus that even when researchers carefully pick instruments that are plausibly exogenous, it is still unlikely that an instrument perfectly satisfies the orthogonality condition. Nevertheless, applied researchers lack formal methods for drawing inferences when their instruments are nearly exogenous.

Existing methods enable researchers to make adjustments for the consistency of their estimates, but do not enable researchers to adjust their inferences. For example, in an influential paper, Acemoglu et al. (2008) use instruments to identify the impact of democracy on national income. Acemoglu et al. (2008) note that their instrumental variable estimates identify an upper bound because there is a small positive correlation between their instrument and the structural error term. However, Acemoglu et al cannot say how this small violation of orthogonality affects inference. This is because standard instrumental variable methods that use t-statistics do not adjust the standard error of the endogenous regressor to account for correlations between the instrument and structural error term.

We introduce a method that enables applied researchers to make reliable inferences in a world where instruments are nearly exogenous and not necessarily perfectly exogenous. Our test accounts for the strength of the violation of the orthogonality condition. In the weak instruments literature, it has been shown that inferring causality becomes more difficult as the instrument becomes weaker. We find that identifying casual effects becomes more difficult as the endogeneity of the instrument increases. Our method enables researchers to make inferences that account for both endogeneity and weakness.

We employ a violation of the exogeneity assumption that allows for a local to zero correlation between the instruments and the structural error. Clearly, our method assumes that instruments exhibit asymptotic orthogonality. However, we also show that our method holds in finite samples even when the correlation between the error and the instrument is constant and nonzero.

Our test is a variant of the Anderson-Rubin (1949) test. We derive the limit of the AR test and show that the limit depends on the correlation between the instruments and structural error term. Furthermore, in larger samples, using critical values for the AR test based on the perfect exogeneity assumption creates massive size distortions when there is, in reality, a mild correlation between the instruments and structural error term.

To deal with these size distortions we propose a novel resampling technique for the AR test. This technique is based on the jackknife histogram estimator in section 2 of Wu (1990). Because the AR test can be written in terms of the sample mean (see equation (4)), we can modify the results in section 2 of Wu (1990). Section 2 in Wu (1990) uses sampling without replacement from the original sample by drawing a fraction of the sample size that is proportional to the full sample. In this setup, we propose a fractionally resampled version of the AR test. We show that choosing half the sample as the block size provides valid but conservative critical values. We also find the asymptotic size of fractionally resampled AR test at half of the sample. This is the limit rejection probability, uniformly over violations of exogeneity parameter, of standard AR test compared with fractionally resampled critical values at half sample. We also show that the asymptotic size is smaller than the nominal size in that scenario. As a competing technique we show that subsampling is oversized. We also conjecture that the Kleibergen (2002) test may not be amenable to the resampling technique that we use. Simulations are conducted to check for the size properties and power of this FAR test. We find that a simple finite sample correction for the block size provides very good power.

Guggenberger (2011) analyzes several tests under local violation of exogeneity, and shows all of them have size problems. However, he finds that the Anderson-Rubin type of test has the best size properties. In related work, Kraay (2009) and Conley, Hansen, and Rossi (2007) both use a Bayesian approach for solving the problem of working with instruments that do not perfectly satisfy the orthogonality condition. They clearly show that even a small violation of the orthogonality condition can lead to entirely different outcomes. When they allow there to be mild violation of exogeneity, they find that the confidence intervals for structural parameters are larger. Conley, Hansen, and Rossi (2007) analyze the support of the correlation parameter. Then they find the confidence interval for the structural parameters given the correlation parameter, and take the union of these intervals. This method provides a conservative solution. Conley, Hansen, and Rossi (2007) also use a local to zero approach: here, they assume the correlation parameter comes from a normal distribution and they characterize its asymptotics. In a third approach, Conley et al. (2007) attach Bayesian priors to this parameter and derive the posterior distribution. Kraay (2009) takes a similar approach to this problem; however, his prior for the correlation parameter is not drawn from a normal distribution. In contrast to these methods that place priors on the correlation between the instrument and structural error term, our method is completely data dependent.

Section 2 describes the problem of making inferences with instruments that violate the exogeneity assumption and we develop a novel way of resampling the AR test. Section

3 considers the asymptotic size of AR test as well as the FAR test. Section 4 considers subsampling and then shows that it will be oversized and will not solve the problem of drawing reliable inferences with instruments violating the exogeneity assumption. Section 4 also contains an analysis of some of the variants of subsampling. Section 5 contains Monte Carlo simulations. Section 6 concludes.

2 Inference and Violations of Exogeneity

We analyze a model that contains a specific violation of the exogeneity assumption. Similar assumptions about the violation of exogeneity have been used by Newey (1985) and Hall and Inoue (2003). The assumption in our model allows for a local to zero covariance between the instruments and the structural error term, and is more flexible than the knife-edged exogeneity assumption used in the instrumental variables estimation literature. The model that we use is:

$$y = Y\theta_0 + u, \tag{1}$$

$$Y = Z\Pi + V, \tag{2}$$

where $cov(u, V) \neq 0$, $Y : n \times m$, $Z : n \times k$, $k \geq m$, we also assume $EZ_iV_i' = 0$, for $i = 1, \cdots n$. Then we assume, for $i = 1, \cdots n$,

$$EZ_iu_i = \frac{C}{\sqrt{n}},$$

where C is a $k \times 1$ vector, each component of that vector C_j ($j = 1, \cdots k$) is a constant and is in a compact subset (Γ) of R^k, $C \in \Gamma \subset R^k$. This assumption allows for a mild correlation between the instruments and the structural error. The correlation can be negative or positive for different instruments. Note that the number of instruments k is fixed (i.e. it does not grow with the sample size). So in a two instrument case, we may have $C = (1, -2)'$. We assume without losing any generality that the lower and upper bounds for the compact subset Γ are the same: (for each $j = 1, 2 \cdots k$) γ_l, γ_u, respectively.

This is the same near exogeneity assumption used by Berkowitz, Caner, and Fang (2008) in which the covariance between the structural error and the instruments is "$C/\sqrt{n}$" and where C is a constant vector.

Note that there are no exogenous control variables in the system. In order to simplify the notation, control variables are not included in (1)(2). Control variables can be projected out to get the same results as here. Our goal is to test $H_0 : \theta = \theta_0$. The Anderson-Rubin (1949) test is used to test H_0 and is described as follows:

$$AR(\theta_0) = [(y - Y\theta_0)'Z/n^{1/2}]\hat{\Omega}^{-1}[Z'(y - Y\theta_0)/n^{1/2}], \tag{3}$$

where $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} Z_iZ_i'u_i^2$.

The Anderson-Rubin test (AR (θ_0)) can also be rewritten as

$$AR(\theta_0) = \bar{S}_n'(var\bar{S}_n)^{-1}\bar{S}_n = n\bar{S}_n'\hat{\Omega}^{-1}\bar{S}_n, \tag{4}$$

where $\bar{S}_n = \frac{\sum_{i=1}^n Z_i u_i}{n} = \frac{Z'(y-Y\theta_0)}{n}$, and $var\bar{S}_n = \hat{\Omega}/n$. We can also demean Z_i in the variance formula and this does not change the asymptotics.

2.1 Assumptions

In this section we introduce our assumptions and discuss them.

Assumption 1. *For $i = 1\cdots, n$, assume $\{u_i, Z_i, V_i\}$ are independent, and*

(i).

$$EZ_i u_i = \frac{C}{\sqrt{n}},$$

where C is $k \times 1$ vector. Note that k is fixed. We allow for both positive and negative covariance at the same in the C vector. $C \in \Gamma$ which is a compact subset of R^k. Without losing any generality assume for each $j = 1, \cdots k$, the bounds of Γ are the same: γ_l is the lower bound (scalar) and the upper bound is γ_u (scalar).

(ii).

$$\Omega = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n EZ_i Z_i' u_i^2,$$

where Ω is positive definite and finite and we assume also

$$\hat{\Omega} \to \Omega \quad a.s.$$

(iii).

$$EZ_i V_i' = 0.$$

Assumption 2.

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^n E\|Z_i u_i\|^3 < \infty.$$

Assumption 1 allows for a small covariance between the instruments and the structural error: it is how we operationalize mild violations of exogeneity. Assumption 1 is discussed above in this section. Assumption 2 is needed for the strong law of large numbers approximation for obtaining the Berry-Esseen bounds. This assumption is discussed in Zhao, Wang, and Wu (2004), as Remark 3 after Corollary 1 in their paper. This is a sufficient condition for the Berry-Esseen bound for the independent case. The triangular array case can also be obtained by Theorem 1 in Zhao, Wang, and Wu (2004), and Assumption 2 is again sufficient. The Berry-Esseen bound is used to prove resampling Central Limit Theorem type of result as (28)(29).

Assumption 2 also provides the sufficient condition in Theorem 23.12 of Davidson (1994) which is sufficient condition for Lindeberg Central Limit Theorem.

2.2 Full Sample Result

In this subsection we derive the limiting distribution of the full sample Anderson-Rubin (1949) test under our violation of the exogeneity condition in Assumption 1. In section 3, we consider also the asymptotic size of $AR(\theta_0)$ test where we look at the rejection probability of $AR(\theta_0)$ uniformly over C, compared to standard χ^2 critical values. To understand the remarks after Theorem 1, note that the limit of $n^{-1/2}\sum_{i=1}^{n} Z_i u_i - EZ_i u_i$ is the zero mean normal random variable/vector L with variance Ω.

Lemma 1. *Under Assumptions 1 and 2, we have*

$$AR(\theta_0) \xrightarrow{d} \chi^2_{k,\frac{C'\Omega^{-1}C}{2}}, \tag{5}$$

where C is the vector indicated in Assumption 1i and the limit in Lemma 1 is a non-central χ^2 distribution with k degrees of freedom and $\frac{C'\Omega^{-1}C}{2}$ as the noncentrality parameter.

Note that the noncentral χ^2 limit in Lemma 1a is explicitly written as

$$\begin{aligned}(L+C)'\Omega^{-1}(L+C) &\equiv L'\Omega^{-1}L + 2C'\Omega^{-1}L + C'\Omega^{-1}C \\ &\equiv \chi^2_k + 2C'\Omega^{-1}L + C'\Omega^{-1}C, \end{aligned} \tag{6}$$

by (26)(27).

Lemma 1 shows that if we use the standard χ^2_k critical values when there is a violation of the exogeneity assumption, then the $AR(\theta_0)$ test will be oversized. This is also what we observe in the simulations in Table 1 for the setups discussed in Section 5. This issue will become clear in section 3, when we explicitly show the asymptotic size of $AR(\theta_0)$.

These findings are related to a recent paper by Guggenberger (2011), which analyzes various identification robust tests when there is a local to zero violation of exogeneity. Guggenberger (2011) finds that the $AR(\theta_0)$ test has the best finite sample properties. Caner (2009) analyzes the $AR(\theta_0)$ test in a many weak moments setup that allows for violations of exogeneity. He finds that when the number of violations is limited, the $AR(\theta_0)$ test does not have size problems.

2.3 Resampling Technique

This subsection contains a description of Wu's (1990) jackknife histogram estimator. We show that this resampling method is useful for recovering the limit for tests of the population mean.

It is well known that in large samples, to test $H_0 : \mu = \mu_0$, where μ is the population mean, and μ_0 is the true value of the mean,

$$\frac{\bar{X} - \mu_0}{\sigma/n^{1/2}} \xrightarrow{d} N(0,1), \tag{7}$$

where $\bar{X}$ is the sample mean out of n observations, and σ is the standard deviation of individual x_i.

We first describe Wu's (1990) resampling technique. We take a subset of size b (block size) from n observations. We resample from data x where $x = (x_1, \cdots x_n)$. The blocks in this resampling from x are x_b with size b, and equal probability of $\begin{pmatrix} n \\ b \end{pmatrix}^{-1}$. This is done via simple random sampling without replacement from the population. The size of the blocks plays a crucial role in our results. Denote this resampling technique by "$*$". Notation such as P_*, E_* refer to calculations under $*$, whereas P, E refer to original probability measure, and the expectation respectively.

The jackknife histogram estimator is basically the resampled counterpart of the simple t test in (7) (see also p.1440, Wu, 1990). So instead of μ_0 we use $\bar{X}$; and, instead of $\bar{X}$ we use $\bar{X}_b$ where this is the sample mean from b observations drawn without replacement from the sample of n observations. Next instead of $\sigma/n^{1/2}$ we use $[(1-f)\hat{\sigma}^2/b]^{1/2}$, where f is a fraction of the sample, $0 < f_l < f < f_u < 1$, f_l is the lower bound, and f_u is the upper bound for fractions. So the term $(1-f)\hat{\sigma}^2/b$ is the variance of $\bar{X}_b$ under simple random sampling without replacement (Cochran, 1977), and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Next, Wu (1990) shows that under simple regularity conditions

$$\sup_t |P_* \left(\frac{\sqrt{b}(\bar{X}_b - \bar{X})}{((1-f)\hat{\sigma}^2)^{1/2}} \leq t \right) - \phi(t)| \to 0 \quad a.s.,$$

where $\phi(t)$ represents the cumulative distribution function of standard normal law. So the cumulative distribution of the jackknife histogram estimator converges to the cumulative distribution of standard normal law.

2.4 Fractionally Resampled Anderson-Rubin Test

For our setup, we are interested in resampling from the following quantity: $Z'u = \sum_{i=1}^{n} Z_i u_i$, where $u = y - Y\theta_0$. Denoting the sample average by $\bar{S}_n = n^{-1}\sum_{i=1} Z_i u_i$, $\bar{S}_b$ is the mean of the simple random sample of size b drawn without replacement from n observations (the mean of the b observations that are drawn out of the sample of n observations). Thus, for block size $b = fn, 0 < f_l < f < f_u < 1$ [1]. Note that we are not directly using the estimators in equations (1.2) and (1.3) of Wu (1990). Instead we benefit from section 2, equations (2.2)(2.3) of Wu (1990). We also extend his case to independent random variables for theoretical contribution and the extension to triangular arrays is simple and is discussed after Theorem 1. We discuss our choice of block size after Corollary 1. Note that the researcher chooses the fraction of the sample f. The fractionally resampled Anderson-Rubin test (FAR (θ_0)) is described as:

$$FAR(\theta_0) = \bar{S}_b'(var_* \bar{S}_b)^{-1}\bar{S}_b = \frac{b\bar{S}_b'\hat{\Omega}^{-1}\bar{S}_b}{(1-f)}, \tag{8}$$

where immediately after equation (2.2) on p.1440 of Wu (1990) or Theorem 2.2. of Cochran (1977) it is shown that $var_*\bar{S}_b = \frac{1-f}{b}\hat{\Omega}$. Observe that the right-hand side in (8) is slightly different than the right hand side in (4). This is due to the property of $var_*\bar{S}_b = \left[\frac{\hat{\Omega}}{b}\right](1-f)$. This will play an important role in the derivation of our main result. The limit for $FAR(\theta_0)$ still does not recover the limit in Lemma 1. Theorem 1 provides a very strong direction and intuition about how to obtain this limit by using a variant of resampling technique. After Theorem 1 we discuss how we provide a valid but slightly conservative limit compared to the limits in Lemma 1. Before deriving the limit for $FAR(\theta_0)$, we define a multivariate normal distribution $L \equiv N(0, \Omega)$ as in Lemma 1. Note that in section 3, we also find the asymptotic size of fractionally resampled Anderson Rubin test.

Theorem 1.*For $0 < f_l < f < f_u < 1$, and the test in (8), define*

$$J_b(t) = P_*(FAR(\theta_0) \leq t).$$

Then, under Assumptions 1 and 2, given (5),

$$\sup_t |J_b(t) - \phi_{mf}(t)| \to 0 \quad a.s.,$$

where $\phi_{mf}(t)$ is the cumulative distribution function of following distribution:

$$(1 + \frac{\sqrt{f}}{\sqrt{1-f}})^2 \chi^2_{k,\frac{f}{1+2\sqrt{f}\sqrt{1-f}}\frac{C'\Omega^{-1}C}{2}},$$

where $\chi^2_{k,\frac{f}{1+2\sqrt{f}\sqrt{1-f}}\frac{C'\Omega^{-1}C}{2}}$ is the non central χ^2 limit with k degrees of freedom and the noncentrality parameter $\frac{f}{1+2\sqrt{f}\sqrt{1-f}}\frac{C'\Omega^{-1}C}{2}$.

[1] We could set $b = [fn]$, where $b/n \to f$ and [.] is the integer part of the number fn. We do not do this because it would make our notation unnecessarily complicated. The notation we employ follows the notation in section 2 of Wu (1990).

Remarks.

1. Note that the limit in Theorem 1 is written explicitly as

$$\chi_k^2(1+\sqrt{\frac{f}{1-f}})^2 + 2\left(1+\sqrt{\frac{f}{1-f}}\right)\sqrt{\frac{f}{1-f}}\,(C'\Omega^{-1}L) + \frac{f}{1-f}\,(C'\Omega^{-1}C)\,, \tag{9}$$

where χ_k^2 is the central χ^2 distribution with k degrees of freedom, and $L \equiv N(0,\Omega)$. This is shown in detail in the proof of Theorem 1 through (34)(35).

2. An important point is to have inference by resampling. It seems that a fixed f will not match the limit in Theorem 1 with the one in Lemma 1 perfectly. But with $f = 1/2$ we get a valid solution pointwise. This is conservative but corrects the size distortion. When the fraction is 1/2, the limit in Theorem 1, equation (9) becomes

$$4\chi_k^2 + 4C'\Omega^{-1}L + C'\Omega^{-1}C. \tag{10}$$

In Lemma 1, the limit is by (6)

$$\chi_k^2 + 2C'\Omega^{-1}L + C'\Omega^{-1}C. \tag{11}$$

Comparing (10) with (11) and since both the $AR(\theta_0), FAR(\theta_0)$ use right hand side critical values, then using the critical values from (10) prevent size distortions. However, this may become too conservative. So for the finite samples, we suggest the following: $f_n \to 1/2$, where f_n is a deterministic sequence. Specifically, we can choose $f_n = 1/2 - \kappa_n$, where $\kappa_n > 0$, for each n and $\kappa_n \to 0$ when $n \to \infty$, κ_n is a deterministic sequence, not a random sequence. Since $f_n \to 1/2$, the limit by using this fraction will be again (10). This can also be seen by following the proof of Theorem 1. But in small samples since $f_n < 1/2$, it may provide some power. We will consider two choices for κ_n in the simulation section.

3. The proof uses Assumption 2 which is for independent data. Triangular arrays can also easily be used, and this is discussed after Assumption 2.

4. In section 3 below, we also obtain the rejection probability of $AR(\theta_0)$ compared with the critical values obtained from the fractionally resampled test ($f = 1/2$), uniformly over C. There we show that asymptotic size is less than the nominal size.

2.5 Power Issues of $AR(\theta_0), FAR(\theta_0)$

In order to simplify the notation and to show the effects of other fractions, the analysis of the power of the $FAR(\theta_0)$ will be conducted with a fixed choice of f. From that we can infer what may happen also at $f = 1/2$ specifically.

Here we briefly show that the $AR(\theta_0)$ is consistent against fixed alternatives when there is strong identification and a mild violation of exogeneity. So in equation (2), assume Π has full column rank, but Assumption 1 still holds. If the true parameter value is θ_1 (for the power exercise here), and $\theta_1 \neq \theta_0, \theta_1 - \theta_0 = l$, where l is a nonzero constant, we test $H_0 : \theta = \theta_0$. Then using equations (2) and (3) with $\theta_1 - \theta_0 = l \neq 0$ it follows that $AR(\theta_0) \xrightarrow{p} \infty$.

The key issue is whether the resampling technique that we suggest will have power, and whether this is consistent against fixed alternatives. The key term in the analysis is $\bar{S}_b$ in equation (8). Note that $(.)_*$ represents resampling b observations out of n without replacement. So we can express

$$\bar{S}_b = \frac{\sum_{i=1}^{b}(Z_i Z_i')_*}{\sqrt{b}}\Pi(\theta_1 - \theta_0) + \frac{\sum_{i=1}^{b}(Z_i V_i')_*}{\sqrt{b}}(\theta_1 - \theta_0) + \frac{\sum_{i=1}^{b}(Z_i u_i)_*}{\sqrt{b}}. \tag{12}$$

Note that the first term on the right hand side can be rewritten as

$$\sqrt{b}\left(\frac{\sum_{i=1}^{b}(Z_i Z_i')_*}{b}\right).$$

And, this term converges in probability to infinity at rate $\sqrt{b} = \sqrt{n}\sqrt{f}$. The test will not be consistent if we use the resampled critical values, even though the test diverges at rate n. However, in finite samples we may gain power with a smaller f. So the simulation exercises are crucial for choosing the correct block size in small samples. And, since the main problem is size, a reasonable power loss is acceptable.

2.6 The Algorithm

Next we write the algorithm to test the null of $H_0 : \theta = \theta_0$ by using the critical values obtained from the empirical distribution function of the half-sampled $FAR(\theta_0)$.

Step 1: First calculate the terms $\hat{\Omega}$ from the full sample of Z_i, u_i as described at the beginning of this section.

Step 2: Denote y_{b_o}, Y_{b_o}, Z_{b_o} as draws of block size $b_o = n/2$ from full sample y, Y, Z without replacement, respectively. Note that $y_{b_o} : n/2 \times 1, Y_{b_o} : n/2 \times m, Z_{b_o} : n/2 \times k$. Form

$$\bar{S}_{b_o} = [Z_{b_o}'(y_{b_o} - Y_{b_o}\theta_0)]_*/(n/2).$$

Step 3. Form $FAR_o(\theta_0)$ by using steps 1-2.

Step 4. Repeat steps 2-4, J times. (J may be 1000, or 5000) Then sort J values of what is found in step 4 to form the empirical distribution function.

Step 5. For a 5% test find the 95 percentile of the empirical distribution function in step 4.

Step 6. Reject the null of $H_0 : \theta = \theta_0$, if the full sample $AR(\theta_0)$ as described in equation (3) is larger than the 95th percentile in step 5.

Note that in the above algorithm, $\hat{\Omega}$ is calculated from the full sample. Only the numerator of the test statistic, the score, has to be resampled. The main technical reason for that is shown in the proof of Theorem 1. Basically for any block size b, p.1440 of Wu (1990) or Theorem 2.2 of Cochran (1977) shows that $var_{*}\bar{S}_b = \frac{1-f}{b}\hat{\Omega}$ in our case.

For f_n, we use $f_n = 1/2 - \kappa_n$ instead of $f = 1/2$ in the above algorithm, if the fraction of the sample is not an integer we choose the next highest integer.

3 Asymptotic Size

In this section we find the asymptotic size of $AR(\theta_0)$ test and the fractionally resampled $FAR(\theta_0)$ test at $f = 1/2$, and then discuss the case of $f < 1/2$. We show that asymptotic size of $AR(\theta_0)$ test is larger than its nominal level, and hence it is oversized. However, we also prove that at half of the sample, the asymptotic size of the $FAR(\theta_0)$ test is less than its nominal size. So even in the case of the largest rejection probability (uniformly over C), using fractionally resampled critical values prevents the size distortions when there are violations of exogeneity. As far as we know this is a new result in the literature.

Guggenberger (2011) derives the first result in this literature about asymptotic size of $AR(\theta_0)$ test, and shows that $AR(\theta_0)$ test is oversized when there is violation of exogeneity. This is very valuable to applied researchers. Our first result about $AR(\theta_0)$ test is a special case of his. However, the next result in this paper about asymptotic size of the fractionally resampled ($FAR(\theta_0)$) is new and shows that this resampling technique may contribute to solving the oversize problem of the $AR(\theta_0)$ test.

Note that subsampling, and m out of n bootstrap results of Theorem 1 of Andrews and Guggenberger (2010) do not apply since subsampling and m out of n bootstrap and the fractional resampling assumptions are different. The difference between subsampling and our case is described below, and it is shown that subsampling will be oversized. This is also shown independently in Guggenberger (2011). In the case of m out of n bootstrap, the main assumption is $b^2/n \to 0$, as $n \to \infty, b \to \infty$, which is not covered by our result. The results for asymptotic size will be proved under (u_i, V_i', Z_i') being iid with distribution F_n, where $1 \leq i \leq n$. The data is iid for each $C \in \Gamma$ as in Guggenberger (2011) where he considers the local violation of exogeneity like us. Note that all limit results in Section 2 is true for triangular arrays as well as iid variables discussed in Remark 3 after Theorem 1, and more in detail after the proof of Theorem 1 in Appendix. Proving Theorem 1 for independent data is just to show that we can extend iid proof technique of Wu (1990) to a general setting that may be more useful to applied researchers. Also note that $\|.\|$ represents the Euclidean norm.

3.1 Asymptotic Size of $AR(\theta_0)$ test

First we start with definition of notation about a parameter space Λ_n. Let $Meval(\Omega)$ represent the minimal eigenvalue of the matrix Ω. For $k \geq 1, d \geq 0, \delta > 0, M < \infty$

$\Lambda_n = \Lambda_n(k,d) = \Lambda_n(k,d,\delta,M) =$

$\{F_n \in R^k, E_{F_n} u_i^2 Z_i Z_i' = \Omega$

$E_{F_n} u_i V_i' = corr(u,V)$ for some $\Omega \in R^{k \times k}$ such that

$E_{F_n} Z_i u_i = C/n^{1/2}$, where C is a subset of a compact in R^k,

$\|\Omega^{-1/2} E_{F_n} u_i Z_i\| \leq d/n^{1/2}$

$E_{F_n} Z_i V_i' = 0$, Meval $(\Omega) \geq \delta$, $E_{F_n} \|Z_i u_i\|^3 \leq M\}$.

Given Assumptions 1-2, this means mainly $\|\Omega^{-1/2}C\| \leq d$. This Λ_n is needed to use the Guggenberger (2011) framework for the asymptotic size of $AR(\theta_0)$. To find the asymptotic size, first we need to find the asymptotic null rejection probability of $AR(\theta_0)$ test. Note that vec(.) operator vectorizes the matrix inside the parentheses. In that respect we define the following:

Definition 1.*For a subsequence ω_n of $n \in N$, denote*

$$\{\lambda_{\omega_n,h} = (F_{\omega_n,h})\}_{n \geq 1},$$

for $h = (h_1', vec(h_2)')$ which is a sequence that satisfies

i). $\lambda_{\omega_n,h} \in \Lambda_{\omega_n}$ for all $n \in N$.

ii). $\omega_n^{1/2}(E_{F_{\omega_n}} u_i^2 Z_i Z_i')^{-1/2} E_{F_{\omega_n}} u_i Z_i = h_1$, and $h_1 \in H$, where H is a compact subset of R^k.

iii). $E_{F_{\omega_n}} u_i^2 Z_i Z_i' \to h_2$ as $n \to \infty$, $h_2 \in R^{k \times k}$.

For Λ_{ω_n} to exist for a given h, we need $\|h_1\| \leq d$, and h_2 to be finite and positive definite. Next, the rejection probability of $AR(\theta_0)$ test under the sequence $\lambda_{n,h}$ is

$$P_{\theta_0,\lambda_n}(AR(\theta_0) > c(1-\alpha)),$$

where $c(1-\alpha)$ is the $1-\alpha$ critical value of the test at nominal size α. In AR test the critical values above come from standard χ_k^2 limit. Note that the subscript on the probability, refers to usage of θ_0, and is calculated under $\lambda_{n,h}$, but to save from notation we denote this as λ_n.

Next we derive the asymptotic null rejection probability of $AR(\theta_0)$ test for $m = 1$, as in Guggenberger (2011) without losing any generality. This will be derived under the sequences $\lambda_{n,h}$. We largely benefit from the idea in the proof of Lemma 1 of Guggenberger (2011) here. We derive the asymptotic null rejection probability from Lemma 1 here. Since the limit in Lemma 1 is continuous in the nuisance parameter, we get Lemma 2 below. To see a detailed proof, p.23 of Lemma 1 in Guggenberger (2011) can be seen. Lemma 2 result is a subcase of his, with our h_1 being in a compact subset.

Lemma 2. *The asymptotic null rejection probability of the $AR(\theta_0)$ test of nominal size α under sequence $\lambda_{n,h}$ is given by*

$$P[\Xi_{k,\|h_1\|^2/2} > \chi^2_{k,1-\alpha}],$$

where $\Xi_{k,\|h_1\|^2/2}$ denotes the non-central χ^2 limit with k degrees of freedom, and $\|h_1\|^2/2$ as the non-centrality parameter. $\chi^2_{k,1-\alpha}$ denotes the $1-\alpha$ quantile of standard χ^2 with k degrees of freedom.

To understand the asymptotic size of $AR(\theta_0)$ test better we need to define the following $k \times 1$ vector that shows the bounds of Γ in Assumption 1. First, $\max(.,.)$ operator shows the maximal of two scalars inside the parentheses. So set

$$\Gamma_k = [\max(|\gamma_l|, |\gamma_u|), \cdots, \max(|\gamma_l|, |\gamma_u|)],$$

where γ_l, γ_u are defined in Assumption 1. Then set $k \times 1$ vector $\Gamma^* = \Omega^{-1/2}\Gamma_k$. Define the asymptotic size of AR test as, given $\lambda_n \in \Lambda_n$,

$$Asize(\theta_0) = \limsup_{n\to\infty} \sup_{C\in\Gamma} P_{\theta_0,\lambda_n}(AR(\theta_0) > \chi^2_{k,1-\alpha}).$$

Note that we could have written the asymptotic size definition uniformly over h_1 as well, this is the same thing as above.

Next, using Lemma 2 here, and the proof of Theorem 2 in Guggenberger (2011), we obtain the following.

Lemma 3.*Given the model, $\Lambda_n(k,d)$ definition for some $\delta > 0$, and $M < \infty$, the asymptotic size of $AR(\theta_0)$ test is*

$$Asize(\theta_0) = P[\Xi_{k,\|\Gamma^*\|^2/2} > \chi^2_{k,1-\alpha}],$$

where $\Xi_{k,\|\Gamma^*\|^2/2}$ is the non-central χ^2 distribution with k degrees of freedom and $\|\Gamma^*\|^2/2$ as the non-centrality parameter.

We can also set this as $\|\Gamma^*\|^2 = d^2$ in Λ_n definition above. This is the same result as Theorem 2 in Guggenberger (2011), in ours the covariance vector is in a compact subset of R^k.

3.2 Asymptotic Size of $FAR(\theta_0)$ test

In this subsection, we consider the asymptotic size of the fractionally resampled AR test. This is the asymptotic rejection probability of $AR(\theta_0)$ compared with critical values from $FAR(\theta_0)$ test (at $f = 1/2$) uniformly over C. For each $\lambda_n \in \Lambda_n$ this is written as

$$\limsup_{n\to\infty} \sup_{C\in\Gamma} P_{\theta_0,\lambda_n}(AR(\theta_0) > c_{n,b}(1-\alpha)) = Asize(\theta_0), \tag{13}$$

where $c_{n,b}(1-\alpha)$ is the $1-\alpha$ quantile of the $FAR(\theta_0)$ at $f = 1/2$. Note that the asymptotic size in (13) can be written in terms of h_1, which is used in Definition 1,

$$\limsup_{n\to\infty} \sup_{h_1 \in H} P_{\theta_0,\lambda_n}(AR(\theta_0) > c_{n,b}(1-\alpha)) = Asize(\theta_0).$$

Note that in order to save notation we use λ_n instead of $\lambda_{n,h}$, where we can see $\lambda_{n,h}$ in Definition 1 with usage of n instead of ω_n. As in Lemma 6vi, and p.460 of Andrews and Guggenberger (2010), we can write the asymptotic size in terms of the subsequence ω_n of n

$$\limsup_{n\to\infty} \sup_{h_1 \in H} P_{\theta_0,\lambda_{\omega_n}}(AR_{w_n}(\theta_0) > c_{\omega_n,b_{\omega_n}}(1-\alpha)) = Asize(\theta_0),$$

where $AR_{\omega_n}(\theta_0)$ represents the $AR(\theta_0)$ test by using ω_n as the sample size, and $c_{\omega_n,b_{\omega_n}}(1-\alpha)$ represents the $1-\alpha$ sample quantile of the fractionally resampled AR test (at $f = 1/2$). This last version of the asymptotic size definition is used in the proof.

Before finding the asymptotic size of $FAR(\theta_0)$, we need the subsequence limit of $AR(\theta_0)$. Given (6), Lemma 2 here and definition of h_1, we have

$$AR(\theta_0) \xrightarrow{d} \chi^2_k + 2h_1'z + h_1'h_1 \equiv J_h, \tag{14}$$

where $z \equiv N(0, I_k)$. Given (14), for a subsequence w_n of n

$$AR_{\omega_n}(\theta_0) \xrightarrow{d} J_h, \tag{15}$$

by using the completion of the sequence argument on p.24-25, equations (4.42)-(4.46) of Guggenberger (2011). The limit of the $FAR(\theta_0)$ at $f = 1/2$, is (by (10))

$$4\chi^2_k + 4C'\Omega^{-1}L + C'\Omega^{-1}C.$$

The same limit can be rewritten as in terms of h_1, where $\Omega^{-1/2}C = h_1$,

$$4\chi^2_k + 4h_1'z + h_1'h_1 \equiv J_{b,h}. \tag{16}$$

In the coming part, we show one of the assumptions that is needed to get the asymptotic size of $FAR(\theta_0)$. In that respect, define the empirical distribution function of $FAR_j(\theta_0)$, for each resample $j = 1, \cdots q_n$ as

$$U_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1_{\{FAR_j(\theta_0) \le x\}}, \tag{17}$$

where q_n represents the number of resamples.

Letting $U_{w_n,b_{w_n}}$ denote the (17) evaluated at subsequence, then we will assume

$$E_{\theta_0,\lambda_{\omega_n}} U_{w_n,b_{w_n}}(x) \to J_{b,h}, \tag{18}$$

under λ_{ω_n} for all $x \in R$, where $J_{b,h}$ is defined in (16). The subscript below expectation here shows that it is under λ_{ω_n}, and we use θ_0, there are no estimators.

Note that the assumption (18) is a high level assumption. Here we discuss how it can be obtained under more primitive conditions. First see that by using the iid nature of the data for each $\lambda_n \in \Lambda_n$, which is assumed in this section,

$$
\begin{aligned}
E_{\theta_0,\lambda_n} U_{n,b_n} &= q_n^{-1} \sum_{j=1}^{q_n} P_{\theta_0,\lambda_n}(FAR_j(\theta_0) \leq x) \\
&= P_{\theta_0,\lambda_n}(FAR(\theta_0) \leq x) \to J_{b,h},
\end{aligned} \tag{19}
$$

via (16), and the remark after Theorem 1. Then, the subsequence version of (19) can be proved via following the proof of Theorem 1, by changing the Assumptions 1-2 to a subsequence form. Another possible way to get the subsequence version of (19) is to use the completion of subsequence to sequence argument but this may be lengthy, and will not be discussed further.

We are now ready to prove one of the main results of this paper. This provides the asymptotic size of the fractionally resampled AR test at $f = 1/2$, and also shows that we do not overreject the null hypotheses. The asymptotic size of $FAR(\theta_0)$ is less than its nominal size. The case for $f < 1/2$ is discussed after Theorem. The case of $f > 1/2$ is discussed in the pointwise case, and it is clear from Remarks to Theorem 1, that it will be conservative with respect to $f = 1/2$. Note that Theorem 2 holds under iid data assumption made in this section.

Theorem 2. *Given λ_n, Definition 1 with $h = (h_1', vec(h_2)')$, in conjunction with Assumptions 1ii-iii, 2, with (18), the asymptotic size of $FAR(\theta_0)$ at $f = 1/2$ is*

i).

$$Asize(\theta_0) = \sup_{h_1}[1 - J_h(c_{b,h}(1-\alpha))].$$

ii). $\alpha > Asize(\theta_0)$,

where α is the nominal size of the AR test, $Asize(\theta_0)$ denotes the asymptotic size of $FAR(\theta_0)$ test. J_h is the limit in (14). But the critical values $c_{b,h}$ correspond to the fractionally resampled test at $f = 1/2$. This is the $1-\alpha$ quantile of the limit in (16). Note that Assumptions 1ii-iii, 2 simplify given the iid assumption on data in this section.

Remarks.

1. Note that with $0 < f < 1$, we can still get an expression for $Asize(\theta_0)$ like in Theorem 2i. The critical values will depend on the limit of (9). However obtaining a result like Theorem 2ii is not clear. This will be case by case in terms of f. In Theorem 2 we show that $f = 1/2$ provides conservative results. Now we consider $0 < f < 1/2$ case. If $f < 1/2$, then $f/(1-f) < 1$, so comparing the critical values of (9) with (11) it will not be very clear whether we have oversized test or not. In that case, the noncentrality parameter of the resampled limit will be smaller than the one in full AR limit, however, the other two terms in the limit of resampled limit is stochastically larger than the ones in full AR limit on the right tail of the distribution. So it will be difficult to judge how α relates to $Asize(\theta_0)$. But for $f = 1/2$, it is clear that $\alpha > Asize(\theta_0)$.

To understand this better, take $f = 1/3$. Then $f/(1-f) = 1/2$. So the limit for fractionally resampled AR test in (9) can be written as

$$\chi_k^2[1+\sqrt{\frac{1}{2}}]^2 + 2\left(1+\sqrt{\frac{1}{2}}\right)\sqrt{\frac{1}{2}}C'\Omega^{-1}L + \frac{1}{2}C'\Omega^{-1}C. \tag{20}$$

The limit in (11) which is the full sample AR test distribution is

$$\chi_k^2 + 2C'\Omega^{-1}L + C'\Omega^{-1}C. \tag{21}$$

So clearly the first two terms on (20) stochastically dominate the corresponding first two terms on (21) on the right hand side of the distribution, but the non-centrality parameter of (20) is half of the one in (21). So at the end of day since the critical values are based on these two limits, we cannot get Theorem 2ii result for $f = 1/3$. However we have done some simulations for calculating the asymptotic size for f near 1/2. For example when we set the compact set for correlations to be in [-.3,.3], with $f = 2/5$, the asymptotic size is 9% at 10% nominal level. The simulation for $f = 2/5$ is based on the format of Table 2 in section 5 below, but with $n = 400$.

2. A relevant issue is what happens to Theorem 2, when $f_n \to 1/2$. Following the proof of Theorem 2, this does not change any of the results there.

4 Comparison With Subsampling and Variants

In this section we compare the resampling technique employed in Theorem 1 with subsampling. We use the limit fraction f in the analysis here just to simplify, this is also valid with f_n^*. Note that subsampling the AR(θ_0) test will not work because it will be oversized. The main reason for this is subsampling described by Politis, Romano and Wolf (1999) allows $b \to \infty, n \to \infty, b/n \to 0$. So $f \to 0$ in our exercise.

A simple counterexample can be seen from our results since in that case $b/n \to 0$, as $b \to \infty, n \to \infty$, meaning $f \to 0$. That means the noncentrality parameter in Theorem 1

$$\frac{f}{1+2\sqrt{f}\sqrt{1-f}}\frac{C'\Omega^{-1}C}{2} \to 0,$$

and

$$\left(1+\frac{\sqrt{f}}{\sqrt{1-f}}\right)^2 \to 1.$$

Hence, the subsampled $AR(\theta_0)$ test statistics will converge to the standard χ^2 limit. Clearly, this is oversized. The subsampled limit is stochastically less than the one in (5).

The work of Andrews and Guggenberger (2007) is important for this section because it analyzes cases in which the subsampling approach works and cases in which it fails. Now we illustrate the difference between subsampling and fractional resampling in a simple example used in section 2 of Andrews and Guggenberger (2007) which analyzes a simple boundary problem. The true parameter θ_0 is nonnegative. Assume that X_i is iid with N (0,1), for $i = 1, \cdots, n$. The Maximum Likelihood Estimator (MLE) of θ_0 is $\hat{\theta}_n = max\{\bar{X}_n, 0\}$, and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The distribution of $\hat{\theta}_n$ is

$$\hat{\theta}_n \sim max\{Z_n, 0\}, \quad Z_n \sim N(\theta_0, \frac{1}{n}).$$

Then we subsample ($b_{n,s}/n \to 0$, as $b_{n,s} \to \infty, n \to \infty$), with $b_{n,s} = o(n)$, the subsampled estimator is $\hat{\theta}_{b_{n,s},j} = max\{\bar{X}_{b_{n,s},j}, 0\}$ where $\bar{X}_{b_{n,s},j} = \frac{1}{b_{n,s}}\sum_{i=j}^{j+b_{n,s}-1} X_i$ and the distribution is

$$\hat{\theta}_{b_{n,s},j} \sim max\{Z_{b_{n,s}}, 0\}, \quad Z_{b_{n,s}} \sim N(\theta_0, \frac{1}{b_{n,s}}).$$

It is clear that the distribution of $\hat{\theta}_{b_n,j}$ does not replicate the distribution of $\hat{\theta}_n$. This is thoroughly discussed in Andrews and Guggenberger (2007). The main reason is that the subsample estimator is closer to the boundary of parameter space than the full sample estimator. To see this $var\hat{\theta}_{b_{n,s},j}$ when $Z_n > 0$ is $1/b_{n,s}$, and $var\hat{\theta}_n$ is $1/n$ when $Z_n > 0$. Since $b_{n,s} = o(n)$, $1/b_{n,s}$ is larger than $1/n$,and hence more variable near the boundary.

Using fractional resampling for this specific example, we set the fractions in a more restrictive way. Set $b_n = fn, f \in [1/2, f_u], f_u < 1$, so $b_n = O(n)$. When we use this technique in the case of the former example (with $Z_{b_n} > 0$), the variance of the fractionally resampled estimator is:

$$var\tilde{\theta}_{b_n} = \frac{1-f}{b_n}.$$

To understand this, we use section 2 of Wu (1990). The variance of the resampled mean is $\frac{(1-f)varX_i}{b_n}$, where $varX_i = 1$ in this example. Then note that compared to the original variance and the subsampled ones

$$var\hat{\theta}_{b_{n,s},j} > var\hat{\theta} \geq var\tilde{\theta}_{b_n},$$

since

$$\frac{1}{b_{n,s}} > \frac{1}{n} \geq \frac{(1-f)}{b_n} = \left(\frac{1-f}{f}\right)\frac{1}{n},$$

and $f \in [1/2, f_u], f_u < 1$. At $f = 1/2$ we have the optimal choice and capture the variance. This shows that variability in this technique is less than or equal to the subsampling technique.

5 Simulation

This section describes the small sample properties of the tests proposed in equation (8). We consider the algorithm in section 2.6, and we resample the critical values according to the deterministic fraction sequence $f_n = 1/2 - \kappa_n$. We set $\kappa_n = 1.5/\sqrt{n}$ or $1/\sqrt{n}$. The other choices for κ_n are tried in terms of convergence rates but these two rates provide the best results. Note that with $n = 100, 200$, the choice of $\kappa_n = 1.5/\sqrt{n}$ leads to block sizes of 35, 79 respectively. For $\kappa_n = 1/\sqrt{n}$ with $n = 100, 200$, the block sizes are 40, 86 respectively.

We consider several block sizes and use the setup in section 2, namely

$$y_i = Y_i\theta_0 + u_i,$$

$$Y_i = Z_i\Pi + V_i,$$

for $i = 1, \cdots, n$. The sample size is n and varies between 100 and 200. We consider the case of one instrument and one endogenous regressor, so $k = 1$, $m = 1$ (exact identification). A case with overidentification is also considered, but not reported here because the results are very similar. For the size exercise, Π can take the values of 2 (strong identification), and 0.2 (weak identification). The iid data (Z_i, u_i, V_i) are generated from a joint normal distribution $N(0, \Omega)$ where

$$\Omega = \begin{bmatrix} 1 & cov(Z_i, u_i) & 0 \\ cov(Z_i, u_i) & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}.$$

So $varZ_i = varu_i = varV_i = 1$, $cov(Z_i, V_i) = 0$, $cov(u_i, V_i) = 0.5$. For the size exercise $\theta_0 = 0$, we test $H_0 : \theta = 0$. For the power $\theta_0 = -2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2$.

We consider three setups for the $cov(Z_i, u_i)$ term. The first setup is consistent with Assumption 1

$$cov(Z_i, u_i) = \frac{C}{n^{1/2}}, \tag{22}$$

and C takes the values of 2, 3 and 5. As C becomes larger, endogeneity becomes more problematic. And, the researcher picks a terrible instrument when $C = 5$.

In the second setup we have:

$$cov(Z_i, u_i) = D, \tag{23}$$

where D is a constant and takes on the values 0.2, 0.3, 0.5. In this setup we expect large size distortions to emerge as the sample becomes larger, because the drift D is multiplied with the square root of the block size in the score in the test statistic. We also used negative values for the covariance term, but the results do not change and hence are not reported.

In the third setup, we have something that is more consistent with bounds of Γ approaching infinity (violation of our Assumption 1).

$$cov(Z_i, u_i) = \frac{an^{1/3}}{n^{1/2}}, \tag{24}$$

where a takes the values of 0.25, 0.5 and 1. At $n = 100$, these correspond to covariances (and correlations, since the variances are normalized at 1) of 0.12, 0.23 and 0.46 respectively for $a = 0.25, 0.5, 1$.

For the power exercise we only use strong identification case, since AR test is known to have low power when there is weak identification. In all exercises, we use 1000 iterations.

Table 1 reports the size of the full sample regular $AR(\theta_0)$ test in (3). This is compared with asymptotic critical values for χ_1^2 distribution at 10% level. We report the rejection rates of the true null in Table 1. We see that both in setups 1 and 2 the actual size is very large. In setup 1, at $C = 2$, the size is 66% with $n = 200$. This shows there is a major size distortion problem if we use the $AR(\theta_0)$ test when there is a violation of exogeneity. This can also be seen for t-tests in Berkowitz, Caner and Fang (2008). The size calculations are done for $\Pi = 2$. Simulations for the case when $\Pi = 0.2$ are also done, but the results are not reported because they are very similar to the case where $\Pi = 2$. Another point is that size gets worse as the sample size increases in setup 2. This case sends an important warning to applied researchers who believe that increasing the sample size can correct for size distortions! In fact, as is very clear from setup 3, a larger sample size can also increase size distortions. With $a = 1$, and $n = 100$, the correlation is 0.46, the instrument is poorly selected and there is a huge size distortion.

Tables 2-4 show the size of the test under two possible f_n sequences. With $f_n = 1/2 - 1.5/\sqrt{n}$, at small to moderate correlation levels size is always less than 10% when the sample size is either 100 or 200. In setups 1 and 3, it is notable that when the sample size increases from 100 to 200, the size decreases from 33% to 1.8% at $C = 5$ in Table 2. Also in Table 4, we see that when $a = 1$ with $n = 100$, the size is 19.5% but with $n = 200$ the size decreases to 7%. We also observe with $f_n = 1/2 - 1/\sqrt{n}$, tests are heavily undersized, so this reflects a power problem with that specific choice.

Note that Tables 2-4 assume homoskedastic errors and sets up the test and the resampling in that way. In Tables 5-7, we introduce conditionally heteroskedastic case. For this we borrow a heteroskedastic setup used in Guggenberger and Smith (2005):

$$u*_i = |Z_i|u_i,$$

where u_i^* represents the new structural error. The conditional heteroskedasticity robust AR statistic and FAR statistic are defined in (3) and (8) respectively, where $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' u_i^2$ is White's conditional heteroskedasticity consistent estimator. Tables 5-8 report actual sizes of the conditional heteroskedasticity robust FAR test under three setups when instruments are strong and weak ($\Pi = 2$, $\Pi = 0.2$).

In Tables 5-7, the size of the test is much smaller at $n = 100$ compared with $n = 100$ in the homoskedastic case in Tables 2-4. We can see that at $n = 100$, with $D = 0.5$ in Table 6, the size is 4.1% at 10% nominal level with $\Pi = 2$. This is the case with $f_n = 1/2 - 1.5/\sqrt{n}$. When we use $f_n = 1/2 - 1/\sqrt{n}$, almost uniformly the size is 0. This will create power problems.

Tables 8-10 provide power results for the homoskedastic case. Clearly with $f_n = 1/2 - 1.5/\sqrt{n}$, we have very good power. For example at Table 8, with setup 1, the power is around 89-100% under various alternatives. However, with the choice of $f_n = 1/2 - 1/\sqrt{n}$, the power declines substantially. At setup 1, the power is around 1-47% under several alternatives.

Tables 11-13 show the power under heteroskedasticity. The results are similar to homoskedastic case. The main difference is that when $f_n = 1/2 - 1.5/\sqrt{n}$ in Table 11. The power is low at alternatives -0.5, 0.5. At other alternatives, the power is around 89-100%.

6 Conclusion

Instruments that perfectly satisfy the knife-edge orthogonality assumption are few and far between. This paper shows that it is possible to conduct inference using instrumental variables when there is a mild violation of the exogeneity assumption. We use a novel resampling technique which draws random blocks from the all of the sample without replacement. The fractionally resampled Anderson-Rubin test (FAR(θ_0)) does not overreject the null in large samples when the fraction is half of the sample. Asymptotic size of the new resampled test is also found. We also provide a finite sample correction for the block size that provides good power.

APPENDIX

Proof of Lemma 1.

By (1)(3) we have

$$\begin{aligned} AR(\theta_0) &= \left(\frac{u'Z}{\sqrt{n}}\right)\hat{\Omega}^{-1}\left(\frac{Z'u}{\sqrt{n}}\right) \\ &= [n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i) + n^{-1/2}\sum_{i=1}^{n}EZ_iu_i]'\hat{\Omega}^{-1} \\ &\times [n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i) + n^{-1/2}\sum_{i=1}^{n}EZ_iu_i]. \end{aligned} \tag{25}$$

First by Assumption 1

$$n^{-1/2}\sum_{i=1}^{n}EZ_iu_i = C.$$

Then (25) can be written as

$$AR(\theta_0) = [n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i) + C]'\hat{\Omega}^{-1}[n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i) + C]. \tag{26}$$

Then in (26) we use the Lindeberg Central Limit Theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i) \xrightarrow{d} N(0,\Omega) \equiv L. \tag{27}$$

Then by Assumption 1, (26)(27)

$$AR(\theta_0) \xrightarrow{d} \chi^2_{k,\frac{C'\Omega^{-1}C}{2}}.$$

Q.E.D.

Proof of Theorem 1.

Define as in Wu (1990)

$$J(t) = P_*\left[\frac{\sqrt{b}(\bar{S}_b - \bar{S}_n)}{[(1-f)v\hat{a}rS]^{1/2}} \le t\right]. \tag{28}$$

Then since Ω being finite and nonsingular, via Assumption 2, Corollary 1 of Zhao, Wang, Wu (2004)

$$\sup_t |J(t) - \phi(t)| \to 0, \quad a.s., \tag{29}$$

where $\phi(t)$ is the standard normal distribution.

Rewrite $FAR(\theta_0)$ as, where $v\hat{a}rS = \hat{\Omega}$,

$$FAR(\theta_0) = [\sqrt{b}(\bar{S}_b - E_*\bar{S}_b + E_*\bar{S}_b)]'((1-f)v\hat{a}rS)^{-1}[\sqrt{b}(\bar{S}_b - E_*\bar{S}_b + E_*\bar{S}_b)]. \tag{30}$$

Then note that

$$\begin{aligned}\sqrt{b}E_*\bar{S}_b &= \frac{\sqrt{b}}{\sqrt{n}}\left[n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i)\right] + \frac{\sqrt{b}}{\sqrt{n}}(\sqrt{n}EZ_iu_i) \\ &= \frac{\sqrt{b}}{\sqrt{n}}\left[n^{-1/2}\sum_{i=1}^{n}(Z_iu_i - EZ_iu_i)\right] + \frac{\sqrt{b}}{\sqrt{n}}C. \qquad (31)\end{aligned}$$

Next, using (28)(29)(31) in (30), $\sqrt{b} = \sqrt{f}\sqrt{n}$, by Assumption 1, Lindeberg Central Limit Theorem with $L \equiv N(0, \Omega)$, and since $J_b(t) = P_*(FAR(\theta_0) \le t)$,

$$\sup_t |J_b(t) - \phi_{\chi^2_f}(t)| \to 0, \quad a.s. \qquad (32)$$

In deriving the result above we use in (30)$(E_*\bar{S}_b = \bar{S}_n)$

$$E_*[(\bar{S}_b - E_*\bar{S}_b)E_*\bar{S}_b] = [E_*\bar{S}_b]^2 - [E_*\bar{S}_b]^2 = 0. \qquad (33)$$

In (32), $\phi_{mf}(t)$ is the cumulative distribution for the following limit:

$$\begin{aligned}[L + \frac{\sqrt{f}}{\sqrt{1-f}}L + \frac{\sqrt{f}}{\sqrt{1-f}}C]' &\times \Omega^{-1}[L + \frac{\sqrt{f}}{\sqrt{1-f}}L + \frac{\sqrt{f}}{\sqrt{1-f}}C] \\ &\equiv (1 + \sqrt{\frac{f}{1-f}})^2\chi^2_k \\ &+ 2\left(1 + \sqrt{\frac{f}{1-f}}\right)\sqrt{\frac{f}{1-f}}[C'\Omega^{-1}L] + \frac{f}{1-f}[C'\Omega^{-1}C]. \qquad (34)\end{aligned}$$

Note that by using the left hand side term in (34) we can rewrite the limit in terms of a non-central χ^2 distribution by combining the terms in L and by dividing and multiplying with $(1 + \frac{\sqrt{f}}{\sqrt{1-f}})^2$

$$\begin{aligned}& [L + \frac{\sqrt{f}}{\sqrt{1-f}}L + \frac{\sqrt{f}}{\sqrt{1-f}}C]'\Omega^{-1}[L + \frac{\sqrt{f}}{\sqrt{1-f}}L + \frac{\sqrt{f}}{\sqrt{1-f}}C] \\ \equiv\ & [L(1 + \frac{\sqrt{f}}{\sqrt{1-f}}) + \frac{\sqrt{f}}{\sqrt{1-f}}C]'\Omega^{-1}[L(1 + \frac{\sqrt{f}}{\sqrt{1-f}}) + \frac{\sqrt{f}}{\sqrt{1-f}}C] \\ \equiv\ & (1 + \frac{\sqrt{f}}{\sqrt{1-f}})^2\chi^2_{k, \frac{f}{1+2\sqrt{f}\sqrt{1-f}}\frac{C'\Omega^{-1}C}{2}}. \qquad (35)\end{aligned}$$

In the last equivalence, note that we use

$$\left(\frac{\sqrt{f}}{\sqrt{1-f}}/(1 + \frac{\sqrt{f}}{\sqrt{1-f}})\right)^2 = \frac{f}{1 + 2\sqrt{f}\sqrt{1-f}}.$$

Q.E.D.

Remark. The only difference between the proof of iid case in Wu (1990) and the one here is the Berry-Esseen bounds. The iid case in Wu (1990) is satisfied under finite second moments as well. The extension to triangular arrays in (28)(29) can be done using Theorem 1 of Zhao, Wu and Wang (2004). The Berry-Esseen bounds are for a sample sum from a finite set of independent random variables as described in Zhao, Wu, Wang (2004).

Proof of Theorem 2i. The proof will consist of simple steps. First we assume Assumption E0 in Andrews and Guggenberger (2009b). For all sequences λ_{ω_n} (where $U_{\omega_n,b_{\omega_n}}$ is the subsequence version of U_{n,b_n})

$$U_{\omega_n,b_{\omega_n}} - E_{\theta_0,\lambda_{\omega_n}} U_{\omega_m,b_{\omega_n}} \xrightarrow{p} 0, \tag{36}$$

under λ_{ω_n} for all $x \in R$. Note that (36) is automatically satisfied since we assume in section 3 that all observations are iid for fixed $\lambda_n \in \Lambda_n$.

So given (36)(18) or using the primitives to derive (36)(18) and definition (17) we have

$$U_{\omega_n,b_{\omega_n}}(x) \xrightarrow{p} J_{b,h}, \tag{37}$$

for all x in continuity points of $J_{b,h}$ under λ_{ω_n}, $n \geq 1$.

Next, use the fact that for $\epsilon > 0$

$$J_{b,h}(c_{b,h}(1-\alpha)+\epsilon)) > 1-\alpha,$$

where $c_{b,h}(1-\alpha)$ is the $1-\alpha$ quantile of $J_{b,h}$. Combine (15)(17)(37), and the above result in Lemma 5 of Andrews and Guggenberger (2010) to have

$$P_{\theta_0,\lambda_{\omega_n}}(AR_{\omega_n}(\theta_0) \leq c_{b_{\omega_n},\omega_n}(1-\alpha)) \to J_h(c_{b,h}(1-\alpha)), \tag{38}$$

where $c_{b_{\omega_n},\omega_n}$ are the of the empirical distribution function of fractionally resampled AR test at $f = 1/2$, and they converge to $c_{b,h}$ which is the $(1-\alpha)$ quantile of $J_{b,h}$ in (16)(37).

To see how asymptotic size is derived, we need the following definition from p.460 of Andrews and Guggenberger (2010)

$$\begin{aligned} Asize(\theta_0) &= \lim_{n\to\infty} P_{\theta_0,\lambda_{\omega_n^*}}(AR_{\omega_n^*}(\theta_0) > c_{b_{\omega_n^*},\omega_n^*}(1-\alpha)) \\ &= \limsup_{n\to\infty} \sup_{h_1 \in H} P_{\theta_0,\lambda_{\omega_n}}(AR_{\omega_n}(\theta_0) > c_{b_{\omega_n},\omega_n}(1-\alpha)). \end{aligned}$$

The worst subsequence ω_n^* is the one that provides the largest rejection probability of the test uniformly over h_1. We could have written also sup over C rather than h_1, since

$\Omega^{-1/2}C = h_1$. To get asymptotic size in our case, we use (38) and the definition of the asymptotic size to have

$$\begin{aligned} \sup_{h_1 \in H} P_{\theta_0,\lambda_{\omega_n}}(AR_{\omega_n}(\theta_0) > c_{b_{\omega_n},\omega_n}(1-\alpha)) &\rightarrow \sup_{h_1 \in H}[1 - J_h(c_{b,h}(1-\alpha))] \\ &= Asize(\theta_0), \end{aligned} \tag{39}$$

since h_1 is in a compact.

Proof of Theorem 2ii. Next, since we consider upper (right-hand side) critical values, clearly

$$c_h(1-\alpha) < c_{b,h}(1-\alpha), \tag{40}$$

for all h_1 and $c_h(1-\alpha)$ is the critical value that corresponds to full sample asymptotics (standard) in (14), and $c_{b,h}(1-\alpha)$ quantile of the distribution in (16). Given (40), (39) it is obvious that by the $Asize(\theta_0)$ definition and J_h being the distribution function of the full sample (standard) $AR(\theta_0)$

$$\alpha > Asize(\theta_0).$$

Q.E.D.

Acknowledgements We thank the co-editor Cheng Hsiao, an associate editor, and an anonymous referee for their comments. We are grateful to Patrik Guggenberger for pointing out an error in an earlier version of the paper, and for the help that he is provided to us. We also thank C.R. Rao and C.F.J. Wu for comments.

References

Acemoglu, Daron, S. Johnson, J. A. Robinson, and P. Yared, 2008, Income and democracy. American Economic Review 98, 808-42.

Anderson, T.W., and H. Rubin, 1949, Estimators of the parameters of a single equation in a complete set of stochastic equations. The Annals of Mathematical Statistics 21, 570-582.

Andrews, D.W.K., and P. Guggenberger, 2007, The limit of finite-sample size and a problem with subsampling. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D.W.K., and P. Guggenberger, 2009a, Hybrid and size-corrected subsampling methods. Econometrica 77, 721-762.

Andrews, D.W.K., and P. Guggenberger, 2009b, Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. Econometric Theory 25, 669-710.

Andrews, D.W.K., and P. Guggenberger, 2010, Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26, 426-468.

Berkowitz, D., M. Caner, Y. Fang, 2008, Are nearly exogenous instruments reliable?. Economics Letters 101, 20-23.

Caner, M., 2009, Near exogeneity and weak identification in generalized empirical likelihood estimators: many moment asymptotics. Working Paper, Department of Economics, North Carolina State University.

Card, D., 1995, Using geographic variation in college proximity to estimate the returns to schooling in L.N. Christofides (Ed.). Aspects of Labour Market Behaviour: Essays in honor of John Vanderkamp.University of Toronto Press, 201-221.

Cochran, W.G., 1977, Sampling Techniques, 3rd Ed. John Wiley.

Conley, T., C. Hansen, and P.E. Rossi, 2007, Plausibly Exogenous. Unpublished Manuscript. Graduate School of Business, University of Chicago.

Davidson, J., 1994, Stochastic Limit Theory. Oxford University Press.

Guggenberger, P., 2011, On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Forthcoming, Econometric Theory.

Guggenberger, P., and R.J. Smith, 2005, Generalized empirical likelihood estimators and tests under partial, weak and strong identification. Econometric Theory 21 , 667-710.

Hall, A.R., and A. Inoue, 2003, The large sample behaviour of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114, 361-394.

Kleibergen, F., 2002, Pivotal statistics for testing structural parameters in instrumental variable regression. Econometrica 70, 1781-1803.

Kraay, A., 2009, Instrumental variables regressions with honestly uncertain exclusion restrictions. Unpublished Manuscript. World Bank.

Newey, W.K., 1985, Generalized method of moments specification testing. Journal of Econometrics 29, 229-256.

Politis, D. J. Romano, and M. Wolf, 1999, Subsampling. Springer Verlag.

Stock, J. and J. Wright, 2000, GMM with weak identification. Econometrica 68, 1055-1096.

Wu, C.F.J, 1990, On the asymptotic properties of the jackknife histogram. Annals of Statistics 18, 1438-1452.

Zhao, L.C., C.Q. Wu, and Q. Wang, 2004, Berry-Esseen bound for a sample sum from a finite set of independent random variables. Journal of Theoretical Probability 17, 557-571.

Table 1: Size at 10%, $AR(\theta_0)$ test, $\Pi = 1$

Sample Size	Setup 1			Setup 2			Setup 3		
	$C=2$	$C=3$	$C=5$	$D=0.2$	$D=0.3$	$D=0.5$	$a=0.25$	$a=0.5$	$a=1$
100	65.0	93.0	100.0	64.0	93.0	100.0	32.8	77.9	93.0
200	66.0	93.0	100.0	88.0	99.0	100.0	42.2	91.7	100.0

Note: Setup 1 is explained in (22), Setup 2 is explained in (23). "D" represents the covariance between the instrument and the structural error. Setup 3 and constant "a"is explained in (24).

Table 2: Setup 1, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n=100$		$n=200$		$n=100$		$n=200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$C=2$	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0
$C=3$	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0
$C=5$	33.1	0.1	1.8	0.0	29.5	0.0	1.8	0.0

Note: This is the test statistic in (8) and setup 1 is (22). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 3: Setup 2, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n=100$		$n=200$		$n=100$		$n=200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$D=.2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$D=.3$	1.6	0.0	0.4	0.0	1.2	0.0	0.2	0.0
$D=.5$	29.6	0.1	34.1	0.0	34.5	0.2	35.9	0.0

Note: This is the test statistic in (8) and setup 1 is (23). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 4: Setup 3, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n=100$		$n=200$		$n=100$		$n=200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$a=.25$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$a=.50$	0.1	0.0	0.0	0.0	0.5	0.0	0.0	0.0
$a=1$	19.5	0.0	7.0	0.0	20.5	0.0	6.9	0.0

Note: This is the test statistic in (8) and setup 1 is (24). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 5: Setup 1, Size at 10%, $FAR(\theta_0)$, Heteroskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$C = 2$	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$C = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$C = 5$	3.7	0.0	0.1	0.0	2.8	0.0	0.0	0.0

Note: This is the test statistic in (8) and setup 1 is (22). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 6: Setup 2, Size at 10%, $FAR(\theta_0)$, Heteroskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$D = .2$	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
$D = .3$	1.6	0.0	0.0	0.0	0.2	0.0	0.0	0.0
$D = .5$	4.1	0.0	0.1	0.0	2.5	0.0	0.9	0.0

Note: This is the test statistic in (8) and setup 1 is (23). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 7: Setup 3, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

	$\Pi = 2$				$\Pi = 0.2$			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
Correction Factor $\kappa_n =$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$	$\frac{1.5}{n^{1/2}}$	$\frac{1}{n^{1/2}}$
$a = .25$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$a = .50$	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
$a = 1$	2.0	0.0	0.4	0.0	20.5	2.1	0.1	0.0

Note: This is the test statistic in (8) and setup 1 is (24). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 8: Setup 1, Power, $FAR(\theta_0)$, Homoskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$C = 2$	100.0	100.0	99.6	98.1	92.3	99.2	99.5	99.3
$C = 3$	99.8	99.9	100.0	95.2	96.2	99.7	99.7	99.8
$C = 5$	99.9	99.9	99.9	89.1	98.3	99.4	99.6	99.9
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$C = 2$	47.3	43.4	37.4	3.7	1.1	13.6	23.6	28.2
$C = 3$	44.0	41.8	38.1	2.1	1.4	15.2	23.3	27.2
$C = 5$	43.8	43.0	34.6	0.6	5.8	18.9	25.0	29.0

Note: This is the test statistic in (8) and setup 1 is (22).

Table 9: Setup 2, Power, $FAR(\theta_0)$, Homoskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$D = 0.2$	99.7	99.8	99.7	96.5	95.1	99.3	99.7	100.0
$D = 0.3$	100.0	100.0	99.7	92.0	97.7	99.8	97.7	100.0
$D = 0.5$	99.9	100.0	99.6	69.5	99.5	99.8	100.0	99.9
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$D = 0.2$	46.3	40.7	35.4	1.6	1.7	13.4	24.1	27.5
$D = 0.3$	44.1	43.0	36.4	1.0	3.5	15.7	26.2	30.2
$D = 0.5$	43.2	41.0	34.0	0.0	12.1	22.5	29.0	31.8

Note: This is the test statistic in (8) and setup 2 is (23).

Table 10: Setup 3, Power, $FAR(\theta_0)$, Homoskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$a = 0.25$	100.0	99.9	99.9	97.6	90.2	99.5	99.7	100.0
$a = 0.50$	99.9	99.8	99.9	96.6	95.5	99.3	99.6	99.7
$a = 1.00$	99.8	99.9	99.9	83.2	98.6	99.7	99.7	99.6
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$a = 0.25$	44.9	42.5	37.8	4.8	1.1	10.7	19.8	25.9
$a = 0.50$	42.1	46.2	34.7	1.0	2.5	16.2	24.6	31.8
$a = 1.00$	45.4	40.6	32.7	0.2	7.2	22.6	28.7	32.7

Note: This is the test statistic in (8) and setup 3 is (24).

Table 11: Setup 1, Power, $FAR(\theta_0)$, Heteroskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$C = 2$	99.5	99.9	99.0	30.1	35.7	88.6	97.4	98.1
$C = 3$	99.9	99.7	98.4	17.3	47.6	90.3	96.6	98.9
$C = 5$	100.0	99.6	99.6	2.0	61.8	90.8	96.3	97.3
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$C = 2$	49.0	43.1	21.3	0.0	0.0	1.3	7.2	12.3
$C = 3$	53.3	50.6	27.3	0.0	0.0	1.3	8.2	10.7
$C = 5$	63.1	63.6	34.9	0.0	0.3	2.3	6.9	11.2

Note: This is the test statistic in (8) and setup 1 is (22).

Table 12: Setup 2, Power, $FAR(\theta_0)$, Heteroskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$D = 0.2$	100.0	99.9	98.1	16.4	45.4	90.5	95.8	98.0
$D = 0.3$	99.8	100.0	99.3	6.1	58.0	90.0	96.3	97.9
$D = 0.5$	100.0	100.0	99.4	0.0	69.4	90.6	94.8	96.9
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$D = 0.2$	54.0	48.6	24.4	0.0	0.0	1.4	7.2	11.1
$D = 0.3$	59.3	57.3	32.7	0.0	0.0	1.8	8.1	12.2
$D = 0.5$	77.4	80.2	44.4	0.0	0.3	2.5	6.3	11.4

Note: This is the test statistic in (8) and setup 2 is (23).

Table 13: Setup 3, Power, $FAR(\theta_0)$, Heteroskedastic Case

	$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$a = 0.25$	99.8	99.7	98.8	37.4	30.0	90.4	96.9	98.4
$a = 0.50$	100.0	99.9	99.0	15.1	44.5	91.6	96.2	98.3
$a = 1.00$	100.0	99.9	99.5	0.3	63.5	92.4	95.0	98.4
	$f_n = 1/2 - 1/\sqrt{n}, n = 200$							
$\theta_0 =$	-2	-1.5	-1	-0.5	0.5	1	1.5	2
$a = 0.25$	43.1	44.3	21.5	0.0	0.0	1.7	7.3	12.0
$a = 0.50$	55.4	48.1	22.9	0.0	0.0	2.0	7.8	12.4
$a = 1.00$	72.2	70.0	40.9	0.0	0.0	3.0	5.4	10.1

Note: This is the test statistic in (8) and setup 3 is (24).